Code Summarization: Ranking of the Generators of Summaries according to each Judge for Conciseness and Fluency & Understandability

Human	**CL 7B**	**CL 13B**	**CL 34B**	**GPT 3.5**	**GPT 4**
GPT-3.5 (4.91)	CL 34B (3.33)	GPT-4 (3.13)	GPT-3.5 (4.01)	CL 13B (4.06)	GPT-3.5 (4.98)
GPT-4 (4.80)	CL 7B (3.30)	human (3.01)	GPT-4 (4.00)	GPT-3.5 (4.06)	human (4.80)
human (4.80)	human (3.23)	GPT-3.5 (3.00)	CL 13B (3.99)	GPT-4 (3.98)	GPT-4 (4.73)
CL 13B (4.62)	CL 13B (3.22)	CL 13B (2.98)	CL 34B (3.97)	CL 34B (3.88)	CL 7B (4.37)
CL 34B (4.51)	GPT-3.5 (3.21)	CL 34B (2.97)	CL 7B (3.95)	CL 7B (3.77)	CL 13B (4.36)
CL 7B (4.45)	GPT-4 (3.14)	CL 7B (2.97)	human (3.94)	human (3.70)	CL 34B (4.11)

Table 1: Conciseness: ranking of the generators of summaries according to each judge, including both humans and LLMs.

Human	**CL 7B**	**CL 13B**	**CL 34B**	**GPT 3.5**	**GPT 4**
GPT-4 (4.82)	GPT-4 (4.34)	GPT-3.5 (4.38)	GPT-3.5 (5.00)	GPT-4 (5.00)	GPT-4 (5.00)
CL 34B (4.81)	CL 34B (4.22)	GPT-4 (4.37)	CL 13B (4.99)	GPT-3.5 (4.98)	GPT-3.5 (4.98)
CL 13B (4.68)	GPT-3.5 (4.21)	human (4.35)	GPT-4 (4.99)	CL 34B (4.91)	CL 7B (4.87)
GPT-3.5 (4.67)	CL 7B (4.19)	CL 13B (4.31)	CL 7B (4.94)	CL 13B (4.86)	CL 13B (4.84)
CL 7B (4.40)	CL 13B (4.17)	CL 7B (4.25)	CL 34B (4.91)	CL 7B (4.82)	CL 34B (4.80)
human (3.71)	human (4.17)	CL 34B (4.07)	human (4.87)	human (4.26)	human (4.46)

Table 2: Fluency & Understandability: ranking of the generators of summaries according to each judge, including both humans and LLMs.